Brain-inspired technologies are advancing apace across Europe and are poised to help accelerate the AI revolution

# Neuromorphic technology in Europe

BY taking loose inspiration from the brain, artificial neural network algorithms have made tremendous progress in artificial intelligence. However, to unlock significant gains in terms of novel real-world capabilities, performance and efficiency, a more ambitious step needs to be taken: to develop a new technology that emulates neural computation directly at the hardware level. The NEUROTECH network presents its vision of these 'neuromorphic' technologies and their innovative potential in Europe.

## Efficient *vs.* power-hungry

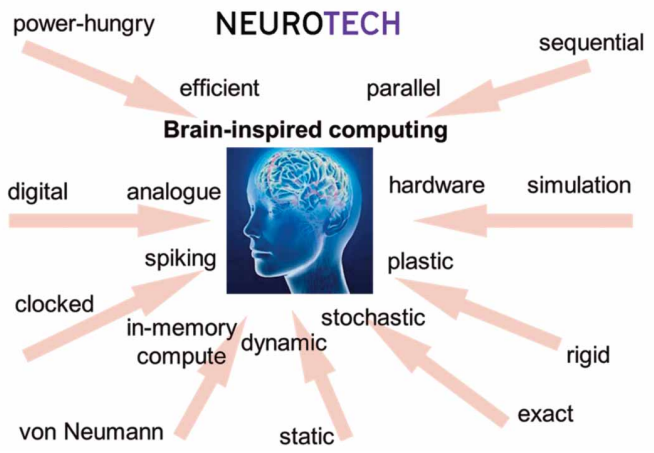
Training artificial neural networks to learn to perform pattern recognition tasks on Graphical Processing Units typically requires hundreds of Watts. Simulating even very small parts of animal brains on supercomputers requires tens of Mega Watts. In comparison, the human brain consumes only 20 Watts to carry out sophisticated perceptual and cognitive tasks. Neuromorphic technologies aspire to emulate neural processing circuits bridging this large energy efficiency gap.

## Parallel *vs.* sequential

Although each neuron typically spikes a few times per second in biological neural processing systems, the massive parallelism of their many neurons and synapses allow them to perform many orders of magnitude more operations per second than those of artificial neural networks simulated on conventional computers. Approaching high levels of parallelism (of the order of thousands and above) in compact and power efficient hardware platforms will require drastic changes in computer architectures and electronic devices.

## In-memory computing *vs.* von Neumann architecture

In conventional computer architectures, a large part of the energy consumption and delays are due to the transfer of information between the physically separated memory and computing parts. In neural network algorithms, this issue ('von Neumann bottleneck') is critical because huge numbers of parameters need to be stored and frequently addressed. Neuromorphic technologies aim at




bringing memory and computing together, like in the brain where computing (neurons) and memory (synapses and topology of the network) are completely intertwined.

## Plastic *vs.* rigid

Learning, both in the brain and in neural networks algorithms, corresponds to repetitive modification of the synapses until reaching a set of connections enabling the network to perform tasks accurately. In conventional computers, this is done by explicit modification of the memory banks storing the weights. Neuromorphic technologies aim at building systems where weights are self-modified through local rules and plastic synaptic devices, as it is done in the brain.

## Analogue *vs.* digital

Conventional computers rely on digital encoding (0 and 1). In the brain, the electrical potential at the membranes of neurons can take continuous values, and so can the synaptic weights. Reproducing such behaviour with digital encoding takes large circuits. Replacing them by analogue components – either CMOS transistors or emerging nanodevices – that directly emulate neural behavior could improve efficiency. However, large scale realisations have yet to be demonstrated.

## Dynamic *vs.* static

Conventional computers use the steady-state of their circuits to encode information. On the contrary, neurons are non-linear oscillators that emit spikes of

voltage. They are coupled to each other and capable of collective behaviour such as synchronisation, transient dynamics and edge of chaos. Neuromorphic technologies aim at emulating such a complex dynamical system in order to go beyond the possibilities of static neural networks, in particular regarding learning.

## Spiking *vs.* clocked

Conventional computers are run by a clock which sets the pace of all circuits. There is no such clock in the brain. In sensory computing, for example, the brain achieves a large part of its efficiency by operating in an event-based manner, where signals are only sampled and transmitted when new information either arrives or is computed. Neuromorphic computing aims at designing spiking architectures natively supporting this scheme.

## Stochastic *vs.* exact

Conventional computers aim at very high precision, contrary to the brain, which neurons and synapses exhibit variability and stochasticity. Resilience to such imprecision seems to be a key asset of neural networks. Relaxing the constraints on the exactitude of components and computing steps in order to decrease energy consumption while maintaining accurate results is a goal of neuromorphic technologies.

Each of these directions represents a breakthrough from the current computing paradigm. As such, neuromorphic computing represents an extremely ambitious multi-disciplinary effort. Each direction will require significant advances in computing theory, architecture and device physics.

## Applications

Neuromorphic computing has the potential to bring huge progress in a wide range of applications. Here we outline some expected important advances.

### Smart Agents on the edge

Neuromorphic technologies will provide systems capable of running state-of-the-art artificial intelligence while consuming little power and energy, enabling embedded and always-on processing. This opens the way to the deployment of artificial intelligence on the edge, where consumption and size are critical. Neuromorphic computing enables continuous learning and adaptation to different environments and users. It shows unprecedented capabilities to encode sensory information efficiently and can lead to smart distributed local processing providing faster responses (to trigger further actions for example) as well as better security and privacy.

### Service to people

Extremely low-power always-on detection systems for voice, speech, and keyword detection can enable further speech processing, towards natural language processing, which in turn will lead to more efficient

“ This opens the way to the deployment of artificial intelligence on the edge, where consumption and size are critical. ”

personal assistants. Always-on systems for fall detection and biomedical signals monitoring people in households will further enhance the capabilities of personal assistants. Robotic assistants will physically interact with the environment and humans and use the adaptability offered by neuromorphic perception and computation to adapt their behaviour. In health, the importance of data privacy is huge, making on-site processing of information even more critical.

### Industrial

Anomaly detection, in its wide sense, is extremely useful in manufacturing plants, where monitoring of workers can improve safety. Neuromorphic technology can also provide solutions for anomaly detection in time series, automatisation of controls and tests, design for manufacturing, defect detection and forecast, predictive maintenance of machines, etc. A similar approach can be deployed in the design of safer cars, both for monitoring the car status and for advanced driver assistance systems, with a limited power budget.

### Research

Besides the impact on the application side, neuromorphic chips will support the research domain, as they are ideal systems to simulate biological neural networks, contributing to understanding the brain and the mechanisms of intelligent behaviour. This would bring massive novel knowledge for new treatments for neurological diseases.

## Technological enablers

Neuromorphic computing requires a departure from the traditional computing paradigm. This implies to use conventional computing substrates in a novel way, or to develop novel substrates. Here we outline key technological enablers for neuromorphic computing, that show promising results.

### Digital CMOS technology

The mainstay of the semiconductor manufacturing industry, digital CMOS is well understood and delivers very consistent performance in volume manufacture. It can access the most advanced semiconductor technologies, which helps offset its intrinsic energy-efficiency disadvantages compared

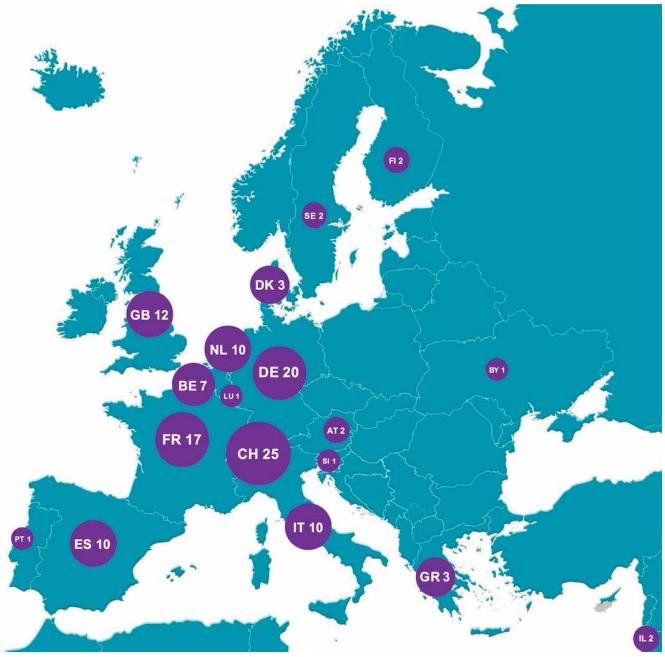
with analogue circuits. When applied to neuromorphic architectures, asynchronous, clocked and hybrid approaches to circuit timing can be used, and algorithms can be mapped into fixed (albeit highly parameterised and configurable) circuits for efficiency or into software for flexibility. Examples of the former include the IBM TrueNorth and Intel Loihi, and of the latter include NEUROTECH partner the University of Manchester's SpiNNaker many-core neuromorphic computing platform.[1]

## Analogue/mixed-signal technology

Event-based analogue mixed-signal neuromorphic technology combines the compact and low power features of analogue circuits with the robustness and low-latency ones of digital event-based asynchronous ones. The key feature of the mixed-signal design approach, compared to pure digital ones, is the ability to build systems able to carry out processing with stringent resources in terms of power and memory by:

- Only dissipating power when the data is present; and
- Processing the data on-line, as it sensed or streamed through the system, using circuits that have time constants matched to the dynamics of the sensory signals processed, and without needing to store data or state variables in memory.

This technology is an enabler for the applications requiring sub-mW always-on real-time processing of sensory signals, for example in edge computing, personalised medicine and Internet of Things domains. Examples of neuromorphic processors that follow this approach are the DYNAP (Dynamic Neuromorphic Asynchronous Processor) series of devices[2] developed by the UZH NEUROTECH members.




## Technologies beyond CMOS

As the CMOS technology approaches its scaling limits, more attention is being devoted to the development of emerging devices, which provide high functionality in a small footprint. In particular, the members of the NEUROTECH network are at the forefront of the development of memristive device technologies, which are a broad class of devices whose resistance can be modified by electrical stimuli.

The leading memristive technologies which are currently at high maturity level are those firstly developed as non-volatile memory devices for storage applications and then integrated in large arrays and in combination with CMOS, namely 'resistive random access memory' (RRAM), 'phase change memory' (PCM), 'ferroelectric memory' (FeRAM), and 'magnetoresistive random access memories' (MRAM). Recently, RRAM, PCM, FeRAM and spin-transfer torque MRAM have been receiving increasing interest for neuromorphic computing, and many hardware demonstrations have been reported at device, but also circuits and systems level. Furthermore, promising developments are underway towards new and less mature concepts which span from new materials (e.g. 2D, nanowires), metal-insulator transition (e.g. $VO_2$-based), organic materials, spintronics (spin torque oscillators, domain walls, spin-waves, skyrmions) and photonics.

## Synapse implementation

The key features of artificial synapses are the ability to update their states given new information (learning, plasticity) and to store analogue information (memory). This can be implemented either with intrinsically analogue or multilevel devices (whether in RRAM, PCM and FeRAM devices, or using magnetic textures such as domain wall or skyrmions), or with binary stochastic devices (as demonstrated for filamentary RRAM, and STT-MRAM). In particular, NEUROTECH members CNR and Un.Zurich have shown how to exploit the non-linear dynamics of analog RRAM synapses to improve the memory lifetime of spiking neural networks based on mixed CMOS-RRAM architecture.[3]

## Neuron implementation

The stochastic, volatile and non-linear properties of memristive device technologies are exploited to emulate neuronal behavior. Among promising technologies, we can mention FeRAM, $VO_2$ –based Metal-Insulator-Transition devices, PCM, STT-MRAM, and spin-torque nano-oscillators. (i.e. specific types of magnetic tunnel junctions, which can be driven into spontaneous microwave oscillations by an injected direct current). In particular, NEUROTECH member CNRS/Thales has shown how to use the non-linear dynamics of the later for processing.[4]

## Challenges for neuromorphic computing

European teams have been at the forefront of developing proofs-of-concept for brain-inspired neuromorphic pattern recognition algorithms (such as the works mentioned above, as well as by NEUROTECH members University of Hertfordshire[5] and University of Heidelberg).[6] It is now time to turn these into concrete low-power and low-latency solutions that outperform conventional approaches. The recently emerged technologies and research directions face four challenges.

### Theoretical foundations

Conventional machine learning draws on linear algebra and calculus, but the theoretical foundations for neuromorphic computing are less clearly defined, as many 'algorithms' behind brain function are only understood qualitatively. Collaborative research between engineers, computer scientists and neuroscientists to translate those findings into hardware and novel substrates can resolve this. Promising algorithms will support theoretically principled supervised and unsupervised learning using only local information.

### Technological maturity

Technologies beyond CMOS transistors in the digital regime suffer from issues like variability in analogue CMOS circuits, lack of endurance in memristive switching devices, difficulty to achieve analogue non-volatile memories, among others. Overcoming these issues requires progress in material and device development, and ideally also novel algorithms that tolerate or even exploit these properties.

### Standardised tools and benchmarks

Developing efficient and effective neuromorphic applications requires knowledge of hardware, network architecture, data representation, and development frameworks. Research has not yet converged on how to best address or standardise any of these.

Moreover, new standard applications, benchmarks and datasets are required since neuromorphic technologies are not necessarily efficient for the same applications as conventional solutions.

### Adoption by industry

The biggest challenge for industry adoption is to find applications which unmistakably demonstrate the potential of neuromorphic technologies, yielding an improvement by at least an order of magnitude over conventional computing approaches. Novel libraries, APIs and GUIs for user-friendly development and debugging are also needed to make neuromorphic systems accessible beyond pure research.

## Neuromorphic computing in Europe and the NEUROTECH network

The European community of neuromorphic computing is extremely active. As of today, over a hundred projects on the subject are funded by the European Union, as shown on the map. Neuromorphic computing is by essence a pluridisciplinary field, requiring many different topics and skills to merge together, such as neuroscience, computer science, electrical engineering, physics, and material science. The emphasis of European funding on collaborative pluridisciplinary projects is a boon for neuromorphic computing, and must be sustained to strengthen this strategic field in Europe. Moreover, it is critical to form an active community around this field, allowing actors from very different backgrounds to come together. The NEUROTECH network aims at promoting neuromorphic computing and forging a European community. NEUROTECH organises events bringing together academics and industry, awards early-career prizes and produces white paper to drive the field.

## References


1 Steve Furber (ed.), Petrut Bogdan (ed.) (2020), 'SpiNNaker: A Spiking Neural Network Architecture', Boston-Delft

2 Moradi *et al*. 'A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)'. *IEEE transactions on biomedical circuits and systems*, 12(1), 106-122. (2018)

3 Brivio *et al*., 'Extended memory lifetime in spiking neural networks employing memristive synapses with nonlinear conductance dynamics', *Nanotechnology* 30(1):015102 (2019)

4 Romera *et al*. 'Vowel recognition with four coupled spin-torque nano-oscillators.'. *Nature* 563.7730: 230-234. (2018)

5 Schmuker *et al*., 'A neuromorphic network for generic multivariate data classification', *PNAS*, 111 (6) 2081-2086, (2014)

6 Schmitt *et al*., 'Neuromorphic hardware in the loop: Training a deep spiking network on the BrainScaleS wafer-scale system', *Proc. IJCNN* (2017)


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824103.


**Co-authors**

E. Donati, M. Payvand, G. Indiveri (Universitat Zurich), P. Debacker (IMEC), S. Furber (University of Manchester), Y. Sandamirskaya (Intel), M. Schmuker (University of Hertfordshire), C. Bartolozzi (Istituto Italiano Di Tecnologia), S. Spiga (Consiglio Nazionale Delle Ricerche), P. Bortolotti (Thales), A. Mizrahi (Thales)

NEUROTECH

**NEUROTECH**

**alice.mizrahi@thalesgroup.com**
**melika@ini.uzh.ch**
**elisa@ini.uzh.ch**
**https://neurotechai.eu**